ACS central science

http://pubs.acs.org/journal/acscii

Research Article

Large-Scale G Protein-Coupled Olfactory Receptor−Ligand Pairing

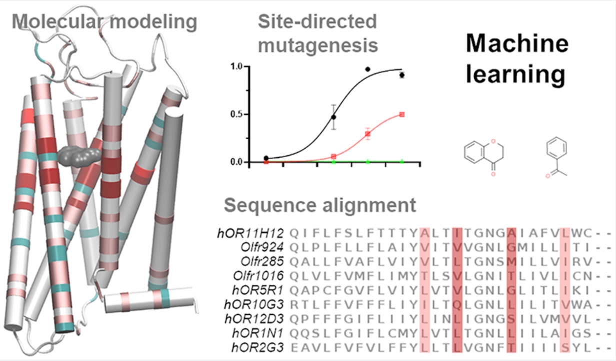
Xiaojing Cong,[*,◆] Wenwen Ren,[◆] Jody Pacalon, Rui Xu, Lun Xu, Xuewen Li, Claire A. de March, Hiroaki Matsunami, Hongmeng Yu, Yiqun Yu,[*] and Jérôme Golebiowski[*]

Cite This: *ACS Cent. Sci.* 2022, 8, 379−387

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: G protein-coupled receptors (GPCRs) conserve common structural folds and activation mechanisms, yet their ligand spectra and functions are highly diverse. This work investigated how the amino-acid sequences of olfactory receptors (ORs)—the largest GPCR family—encode diversified responses to various ligands. We established a proteochemometric (PCM) model based on OR sequence similarities and ligand physicochemical features to predict OR responses to odorants using supervised machine learning. The PCM model was constructed with the aid of site-directed mutagenesis, *in vitro* functional assays, and molecular simulations. We found that the ligand selectivity of the ORs is mostly encoded in the residues up to 8 Å around the orthosteric pocket. Subsequent predictions using Random Forest (RF) showed a hit rate of up to 58%, as assessed by *in vitro* functional assays of 111 ORs and 7 odorants of distinct scaffolds. Sixty-four new OR−odorant pairs were discovered, and 25 ORs were deorphanized here. The best model demonstrated a 56% deorphanization rate. The PCM-RF approach will accelerate OR−odorant mapping and OR deorphanization.


Molecular modeling
Site-directed mutagenesis
1.0
0.5
0.0
Machine learning
Sequence alignment
hOR11H12 QIFLFSLFTTTYALTITGNGAIAFVLWC--
Olfr924 QLPLFLLFLAIYVITVVGNLGMILLITI--
Olfr285 QALLFVAFLVIYVLTLTGNSMILLVIRV--
Olfr1016 QLVLFVMFLIMYTLSVLGNITLIVLICN--
hOR5R1 QAPCFGVFLVIYLVTVLGNLGLITLIKI--
hOR10G3 RTLFFVFFFLIYILTQLGNLLILITVWA--
hOR12D3 QPFFFGIFLIIYLINLIGNGSILVMVVL--
hOR1N1 QQSLFGIFLCMYLVTLTGNLLIIILAIGS--
hOR2G3 EAVLFVFVLFFYLLTLVGNFTIIIISYL--

■ INTRODUCTION

Decoding the sequence−function relationship of proteins is extremely challenging. Slight changes in the sequence may significantly affect the function, whereas proteins with low sequence identity may exhibit similar functions. G protein-coupled receptors (GPCRs) are the most remarkable examples of this phenomenon. They are the largest membrane protein family and the targets for about 40% of marketed drugs.[1] The human genome contains over 800 genes coding for GPCRs,[2] which exert differentiated and specific functions in the complex cellular signaling network. Half of these genes are olfactory receptors (ORs) that endow us with fascinating capacities of odor discrimination.[3] Mammalian GPCRs conserve a typical structure of seven transmembrane helices (7TM) that house an orthosteric ligand-binding pocket.[4] They show a conserved signaling mechanism that involves large-scale conformational changes to accommodate their cognate G proteins. The mechanism is encoded in conserved motifs throughout the 7TM, which form a network of inter-TM contacts converging at the cytoplasmic side.[5] Specifically, the "D(E)RY", "CWLP", and "NPxxY" motifs in TM3, TM6, and TM7, respectively, are the most conserved hubs of the allosteric communication between the orthosteric pocket and the cytoplasmic side of class A GPCRs.[4] The orthosteric pocket, by contrast, has diversified extensively and resulted in huge variations in the receptors' function.

This study focuses on the functional heterogeneity of ORs and how this is encoded in the OR sequences. ORs discriminate a vast spectrum of volatile molecules (odorants) and code for an innumerous number of odors perceived in the brain. The many-to-many relationships between ORs and odorants are key to understanding odor perception.[6] Although odorant-binding proteins (OBPs) also contribute to odor detection, they are abundant extracellular proteins that participate in perireceptor events by selecting/carrying odorants.[7,8] Currently, OR−odorant interactions are mostly measured in heterologous cells, especially for human ORs, which neglects the effect of OBPs. ORs are also expressed ectopically, and some have emerged as appealing drug targets.[9−12] We sought to predict OR responses to various odorants using OR sequence alignment, proteochemometrics (PCM),[13] and machine learning. The PCM model was based on the OR sequence similarities and the chemical features of the odorants. Sequence-based approaches can handle large protein families and circumvent the difficulties in obtaining high-resolution structures, as is the case for ORs. Machine learning models using protein sequences and ligand chemical similarities have shown great success in predicting drug−target interactions, such as reviewed in refs 14−16. Attempts to

Received: December 7, 2021
Published: February 18, 2022

https://doi.org/10.1021/acscentsci.1c01495

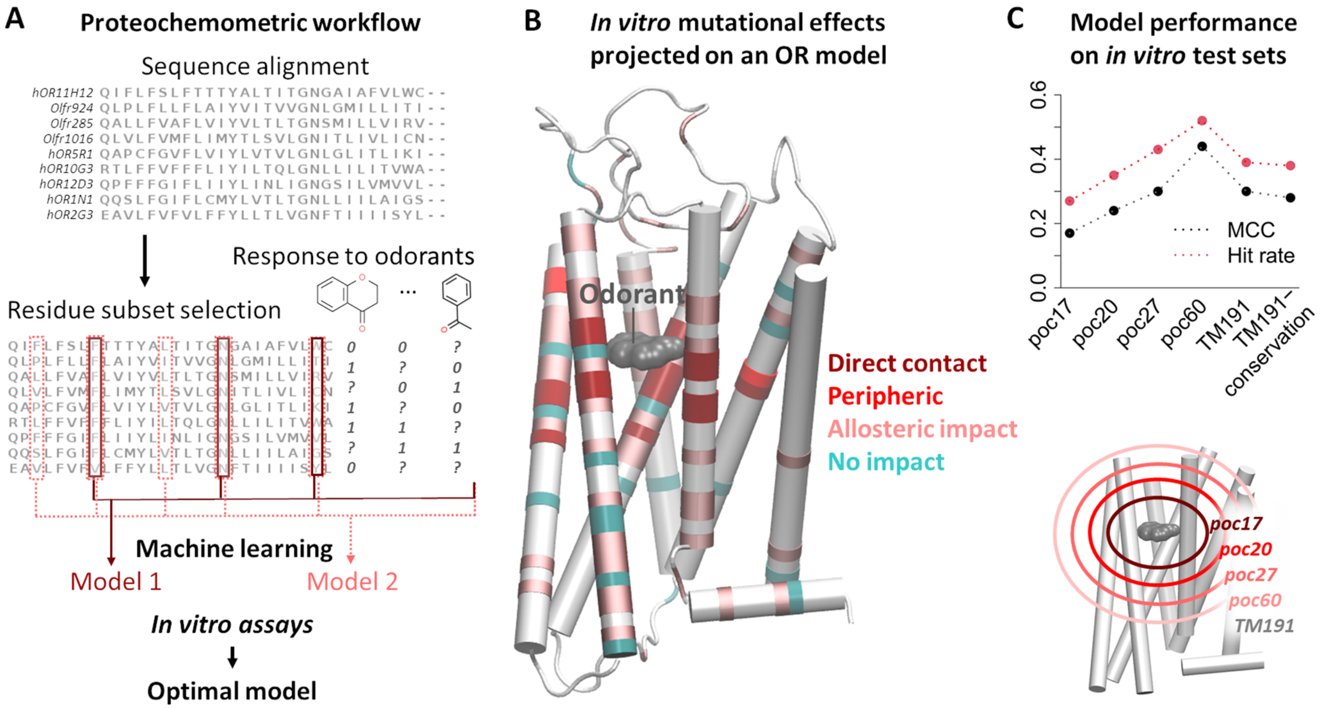

A
Proteochemometric workflow
Sequence alignment
hOR11H12 QIFLFSLFTTTYALTITGNGAIAFVLWC--
Olfr924 QLPLFLLFLAIYVITVVGNLGMILLITI--
Olfr285 QALLFVAFLVIYVLTLTGNSMILLVIRV--
Olfr1016 QLVLFVMFLIMYTLSVLGNITLIVLICN--
hOR5R1 QAPCFGVFLVIYLVTVLGNLGLITLIKI--
hOR10G3 RTLFFVFFFLIYILTQLGNLLILITVWA--
hOR12D3 QPFFFGIFLIIYLINLIGNGSILVMVVL--
hOR1N1 QQSLFGIFLCMYLVTLTGNLLIIILAIGS--
hOR2G3 EAVLFVFVLFFYLLTLVGNFTIIIISYL--
Response to odorants
Residue subset selection
0 0 ?
1 ? 0
? 0 1
1 ? 0
1 1 ?
? 1 1
0 ? ?
Machine learning
Model 1
Model 2
In vitro assays
Optimal model
B
In vitro mutational effects projected on an OR model
Odorant
Direct contact
Peripheric
Allosteric impact
No impact
C
Model performance on in vitro test sets
0.0
0.2
0.4
0.6
MCC
Hit rate
poc17
poc20
poc27
poc60
TM191
TM191–conservation
poc17
poc20
poc27
poc60
TM191

Figure 1. Machine learning protocol and residue selection. (A) Machine learning workflow, in which different residue subsets were extracted from the sequence alignment for the training of different models. The PCM approach combined the OR sequence features, the ligand physicochemical features, and the response data (if available) of each OR–ligand pair. (B) Available site-directed mutagenesis data (including literature data, summarized in ref 24) projected on the 3D model of mOR256-31. Residues in dark red and red belong to *poc17* and *poc20*, respectively. (C) Matthew's correlation coefficient (MCC)[28] and hit rate of the RF classifiers on the *in vitro* test set.

predict OR responses to odorants have also achieved encouraging results.[17−20] However, data scarcity in the immense odor space is a major bottleneck for good predictivities. To date, less than 50% of human ORs (hORs) and 20% of mouse ORs (mORs) have been deorphanized with less than 250 odorants (Table S1). One effective way to handle data scarcity is dimension reduction, such as by selecting relevant residues in the OR sequences (the so-called feature selection). A recent study on insect and mammalian ORs demonstrated that selecting subsets of 20 residues could indeed increase the model predictivity.[20] However, if one assumes that a given function is mostly encoded by 20 residues out of a GPCR sequence of ∼300 residues, the binomial coefficient [300!/20!(300 − 20)!] gives more than 10^{30} possible combinations. Therefore, selecting relevant residues is key to constructing an effectual model.

Like other GPCRs, ORs respond to their ligands via allosteric mechanisms, which involve distinct interwound factors: ligand affinity, intrinsic stability of different receptor states, as well as long-range allosteric coupling between the ligand-binding pocket and the cytoplasmic side.[21] Ligand affinity is thought to be dictated by the residues outlining the binding pocket.[22,23] ORs that respond to the same odorants share higher sequence homology around the pocket than in the rest of the receptor sequence.[18]

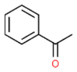
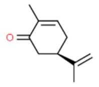
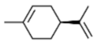
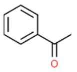
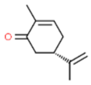
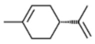
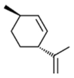
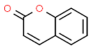
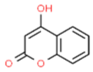
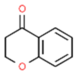
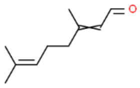
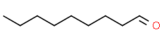
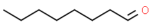
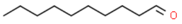
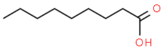
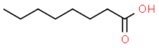
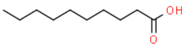
The OR response to odorants can be drastically altered by mutations that are distant from the pocket.[24] It is nontrivial to select the relevant residues. Here, we combined molecular modeling, site-directed mutagenesis with *in vitro* functional assays, and machine learning to identify the most relevant residues. PCM modeling and random forest (RF) were employed to predict OR responses to prototypical odorants using the relevant residues. Finally, *in vitro* functional assays were performed to assess the selection of relevant residues as well as the predictivity of the PCM-RF model. This approach (outlined in Figure 1A) largely outperformed existing models by enabling knowledge-based residue selection. It illustrated how the functional heterogeneity of G protein-coupled ORs is encoded in the sequence.

■ RESULTS

Database of OR–Odorant Pairs for Model Training. We examined all of the literature data of *in vitro* dose-dependent responses of hORs and mORs to diverse odorants. These include 1293 OR–odorant pairs consisting of 390 ORs and 244 odorants. In addition, we included more than 14 400 OR–odorant pairs which have been reported to be non-responsive *in vitro*. The database (Data File S1) contains 720 distinct ORs (including 318 orphan ORs) and 244 odorants. Four odorants were considered here as test cases: acetophenone, coumarin, R-carvone, and 4-chromanone. They have been associated with many ORs (dozens to hundreds) in previous studies (Table 1). To enlarge the training set, we also included the data of 6 additional odorants that have similar chemical structures to the 4 target odorants.

Selection of Relevant Residues. *Molecular Modeling.* Given the existing knowledge of GPCR structures, we first sought for odorant-binding residues within the orthosteric ligand-binding pocket. The mouse OR mOR256-31 (gene name *Olfr263*) was chosen as a prototype, since it is a broadly tuned receptor which responds to three of the four odorants (coumarin, R-carvone, and acetophenone).[25,26] We built a 3D homology model of mOR256-31 bound with the odorants using our previously established approaches and molecular dynamics simulations.[24,25,27] The 3D model was built under

Table 1. Chemical Structure, PubChem CID, and Training Data[a] of the Query Odorants (in Bold) and Their Analogues

Acetophenone, CID: 7410 *P*: 11 hORs, 78 mORs *N*: 183 hORs, 91 mORs		**R-carvone**, CID: 439570 *P*: 7 hORs, 11 mORs *N*: 7 hORs, 46 mORs		S-limonene, CID: 439250 (R-carvone analog) *P*: 4 hORs	
Methyl benzoate, CID: 7150 (acetophenone analog) *P*: 2 hORs, 22 mORs *N*: 255 hORs		S-carvone, CID: 16724 (R-carvone analog) *P*: 6 hORs, 10 mORs *N*: 7 hORs, 42mORs		R-limonene, CID: 440917 (R-carvone analog) *P*: 12 hORs, 1 mOR *N*: 186 hORs, 3 mORs	
Isolimonene, CID: 22831540 (R-carvone analog) *P*: 2 hORs		**Coumarin**, CID: 323 *P*: 7 hORs, 11 mORs *N*: 8 hORs, 51 mORs		4-Hydroxycoumarin, CID: 54682930 (coumarin analog) *P*: 1 mOR *N*: 10 hORs, 51 mORs	
4-chromanone, CID: 68110 *P*: 3 hORs, 13 mORs *N*: 7 hORs, 39 mORs		**Citral**, CID: 638011 *P*: 17 hORs, 1 mOR *N*: 176 hORs, 5 mORs			
Nonanal, CID: 31289 *P*: 9 hORs, 7 mORs *N*: 9 hORs, 47 mORs		Octanal, CID: 454 (nonanal analog) *P*: 4 hORs, 4 mORs *N*: 231 hORs, 52 mORs		Decanal, CID: 8175 (nonanal analog) *P*: 7 hORs, 3 mORs *N*: 85 hORs, 49 mORs	
Nonanoic acid, CID: 8158 *P*: 5 hORs, 12 mORs *N*: 9 hORs, 42 mORs		Octanoic acid, CID: 379 (nonanoic acid analog) *P*: 1 hOR, 9 mORs *N*: 9 hORs, 47 mORs		Decanoic acid, CID: 2969 (nonanoic acid analog) *P*: 2 hORs, 9 mORs *N*: 9 hORs, 43 mORs	

[a] *P*: number of responsive (positive) ORs. *N*: number of nonresponsive (negative) ORs. See Data File S1 for the lists of ORs.

the constraints of conserved amino-acid motifs and site-directed mutagenesis data covering nearly 50% (95 residues) of the TM domain.[24] Seventeen residues were identified within a 5 Å distance of the bound odorants (Table S2). Fourteen of these residues had been shown to be important for OR responses to odorants by site-directed mutagenesis (Table S2). These 17 residues were assumed to be in direct contact with the odorants (named *poc17* hereafter, Figure 1B). However, the relevant residues should include many more than the sole binding pocket.

Site-Directed Mutagenesis. Twenty-four point-mutations were generated within and around *poc17* of mOR256-31. Their impact on the receptor's response to five ligands was measured by *in vitro* dose-dependent responses (Figure S1). We projected the mutational effect onto the 3D model of mOR256-31, together with all of the OR mutations reported in the literature (Figure 1B). Twenty residues including *poc17* and 3 peripheric residues (Figure 1B) delineated a larger orthosteric pocket (*poc20*). Mutations within *poc20* consistently affected the response to most of the odorants. Beyond the region of *poc20*, the mutational effect was less systematic (Figure 1B).

To determine the best subset of residues for predicting OR responses to odorants, we proceeded in an empirical approach. Namely, we selected 5 small-to-large residue subsets as heuristics, based on the above results: *poc17*, *poc20*, *poc27*, *poc60*, and *TM191*. *poc27* and *poc60* are extensions of the pocket until 6 and 8 Å from the bound odorant, containing 27 and 60 residues, respectively (Figure 1C and Table S3). *TM191* contains the whole 7TM region made up of 191 residues. Machine learning models were then built with these residue subsets to compare their predictive power.

PCM and Machine Learning. From the sequence alignment of hORs and mORs, each of the 5 heuristic residue subsets were extracted. PCM models were constructed using the data in Table 1 and physiochemical features of the odorants (see the Material and Methods section). Each OR−odorant pair was labeled with the *in vitro* response (responsive or nonresponsive). We trained and assessed supervised support vector machine (SVM) and RF classifiers using 5-fold cross validation. The response probability of each OR−odorant pair was predicted, and a probability >0.5 was classified as responsive. The predictivity was measured by Matthew's correlation coefficient (MCC).[28] RF performed better than SVM. The predictivities of the five RF classifiers were not significantly different from one another.

However, they were clearly superior to a naive statistical inference (Figure S2A; see the Supplementary Methods section for the calculation of the statistical inference). The *poc60* classifier performed the best on average (Figure S2A, Data File S2A,B). Control models built with 60 randomized residues, as expected, showed no predictivity (Figure S2A). To determine the best residue subset, we constructed five final RF classifiers (*poc17*, *poc20*, *poc27*, *poc60*, and *TM191*) using

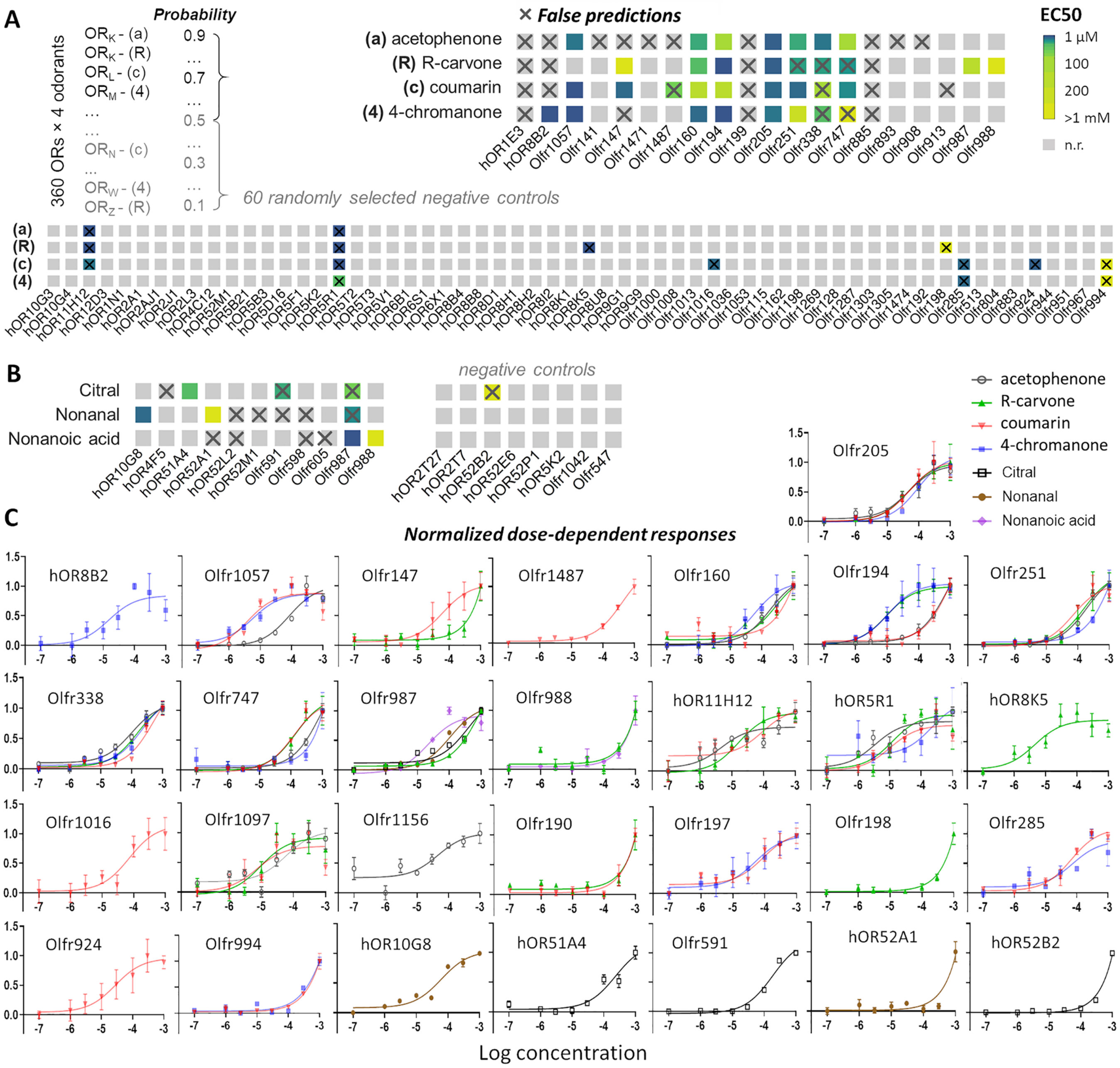


Figure 2. *In vitro* evaluation of machine learning predictions of OR responses to odorants. (A) All of the OR−odorant pairs were ranked by the predicted probability to be responsive. The initial model assessments focused on four odorants. 20 responsive and 60 nonresponsive ORs (negative controls) predicted by the *poc60* model were selected for functional assays. Heatmaps show the *in vitro* EC50 values, in which the false predictions are labeled with ×. Assessments of the other models are provided in Figure S3. (B) *In vitro* assessment of the *poc60* model predictivity for acyclic odorants. (C) Dose-dependent response curves of all of the responsive OR−odorant pairs identified in this study. Error bars indicate SEM (n = 3−6).

100% of the data in Table 1. Each classifier was then used to screen for new ORs for acetophenone, R-carvone, coumarin, and 4-chromanone. The *in silico* screening was performed on 360 ORs (223 hORs and 138 mORs), including 346 orphan ORs. Each classifier predicted and ranked the probabilities of the ORs to respond to each of the 4 odorants (Data File S2C).

In Vitro Assessment of Relevant Residues. We tested the predictions of all five classifiers in cell functional assays. For each model, we tested all ORs in the responsive class (predicted response probability >0.5 for any odorant) as well as 60 negative control ORs (response probability <0.5 for all odorants). These ORs were tested against all 4 odorants. For instance, in the case of *poc60*, we tested all 20 ORs in the responsive class and 60 randomly picked negative controls from the nonresponsive class (Figure 2). Similar tests were performed on the other four models (Figure S3 and Table S4, Data File S2C,D). When significant responses were observed at 300 μM, dose-dependent responses were measured. Otherwise, the OR−odorant pair was considered nonresponsive. The *poc60* classifier performed the best on the *in vitro* test set (Figure 1C). It showed 0.39−0.60 hit rates and 0.43−0.48 predictivity (MCC) for the 4 odorants (Table 2). Therefore, *in vitro* data confirmed that *poc60* is the most relevant residue subset to decode the receptor's response to odorants. These

Table 2. Performance of the *poc60* Model in Predicting New OR−Odorant Pairs[a]

	initial test odorants				additional test odorants		
metrics[b]	acetophenone	R-carvone	coumarin	4-chromanone	citral	nonanal	nonanoic acid
MCC	0.47	0.45	0.43	0.48	0.24	0.48	0.40
hit rate (precision)	0.39	0.6	0.58	0.6	0.50	0.50	0.25
recall (sensitivity)	0.78	0.46	0.47	0.5	0.25	0.67	1.00
F1 score	0.52	0.52	0.52	0.55	0.33	0.57	0.40
specificity	0.85	0.94	0.92	0.94	0.93	0.88	0.65
AUC	0.84	0.72	0.72	0.74	0.58	0.66	0.74
true positives	7	6	7	6	1	2	2
true negatives	60	63	60	64	14	14	11
false positives	11	4	5	4	1	2	6
false negatives	2	7	8	6	3	1	0

[a]See Data File S2C for the raw data. [b]See the Methods section in the SI for the definitions.

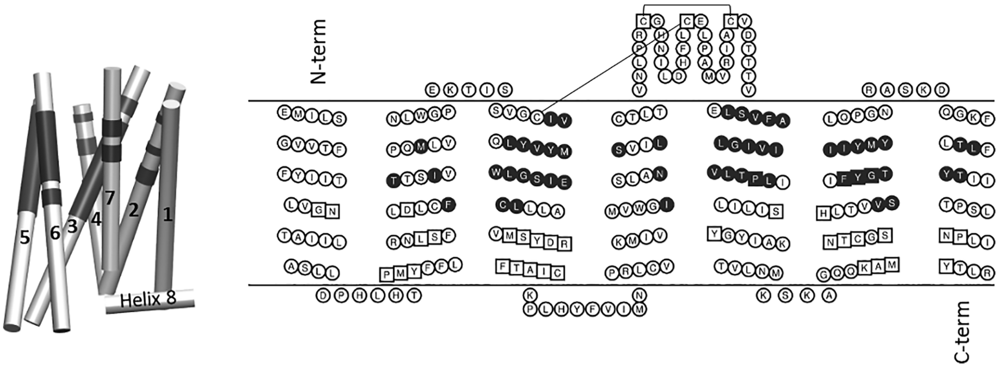


Figure 3. Location of the residues that best encode OR responses to ligands, illustrated with mOR256-31. Conserved motifs in ORs are squared. The N- and C-termini are truncated for clarity.

residues show very low conservation in hORs and mORs (Figure S2B), suggesting that they have diversified to adapt to various ligands.[22,23] This implies that amino acid conservations in the OR sequences contain essential information for their functionality. Thus, we tested an additional model using the amino acid conservations in the TM region. This model turned out to be nearly as predictive as using the amino acid physicochemical features (Figure 1C). This indicates that the type of features used to describe the amino acids is not critical, as long as the features sufficiently convey the sequence differences to the machine learning algorithm.

Assessment of Model Utility. *Applicability to Other Odorants.* While 50% of hORs and 20% of mORs have been deorphanized at the time of this study, only a tiny fraction of the odorant chemical space (<250 odorants) has been tested. The lack of data on odorants is a major restraint on the model utility. To explore this limitation, we generated a learning curve of the *poc60* model predictivity on the external test set versus the amount of training data used (Figure S4A). The learning curve suggested that a meaningful prediction could be obtained for an odorant with ∼15 known ORs. In the current database containing 244 odorants, only 17 (7%) met this criterion, 11 of which contained aromatic or cyclic structures. We attempted three more odorants that contain alkyl chains, citral, nonanal, and nonanoic acid. Following the same procedure, we tested *in vitro* all 11 ORs that were predicted to respond to any of the three odorants as well as 8 negative control ORs (Figure 2B). Because the training data lacked responsive ORs for these odorants, the model predicted less responsive pairs than for the 4 cyclic odorants. *In vitro* assays showed that the model performed well on nonanal and nonanoic acid but not on citral (Table 2). The poor predictivity on citral was likely due to the lack of analogues (thus the lack of data) in the training set (Table 1) and the fact that citral is a mixture of two isomers, which add ambiguity to the available data. The results demonstrate that the model is generalizable to odorants of different chemical groups, provided enough training data for the odorants in question or their close analogues.

General Model Performance. We evaluated the general performance of the *poc60* model on all of the external test set data, including those tested for the other models and for citral. The test set data were shuffled and split into 5 folds, like in a cross validation. The model predictivity was coherent on the 5 folds of the data set, which gave 0.39−0.46 hit rates and 0.32−0.34 MCC (Table S6). Blind OR−odorant screening hit rates in Hana3A cells are expected to be lower than 0.1, such as in a pioneer study on 245 hORs and 219 mORs against 93 odorants.[19] Note that the odorants tested here might be more promiscuous than average, since the model requires training data for the query odorants or their analogues. Our test set also enriched more responsive ORs (26%) than in the natural pool of ORs (e.g., 13% in ref 19), despite the large number of negative-control ORs included. Since many ORs fail to express on the membrane of heterologous cells, it is difficult to estimate the general response rate of ORs to various odorants.

The total external test sets in this work contained 111 ORs and 438 OR−odorant pairs. We identified 63 new OR−odorant pairs with EC50 values in the micromolar to millimolar range, corresponding to 29 ORs (Figure 2C, Figure S3 and Table S5). Twenty-five ORs were deorphanized in this study, including 9 from the negative control groups. Nevertheless, the deorphanization rate is significantly higher in the predicted positive groups than in the negative control groups (Figure S4B), which are 56% and 15%, respectively, for the *poc60* model.

Utility for New ORs and Odorants. One important aspect of the model utility is its predictivity on new ORs and odorants that are not part of the training set. While 56 out of the 95 ORs in the external test set are "new", we recalculated the model performance metrics for this part of the test set. The model still showed good predictivity compared to the full test set (Table S7). The model predictivity on new odorants was evaluated by the following test: we excluded the 7 odorants one by one from the training set, retrained the model, and calculated the performance metrics on the test set containing only the excluded odorant. In this case, the model only showed predictivity for cyclic odorants, acetophenone, R-carvone, and 4-chromanone (Table S8). Therefore, the application to new odorants is currently limited by the lack of training data, as already discussed above. New data will gradually enable the application to more odorants. Currently, the model is readily applicable to new ORs for which there are no training data.

DISCUSSION

This work illustrates how the G protein-coupled ORs' response to ligands can be decoded from their sequence. Sixty residues around the odorant-binding pocket contain the highest signal-to-noise ratio and dictate the variation in the ORs' response to the odorants (Figure 3). The ligand-binding pocket of GPCRs has highly diversified during evolution to discriminate various stimuli. It is not surprising that the ORs' response to the odorants could be predicted by using less than 20% of the sequence, made up with highly variable residues. The results validate previous predictions of pocket residues based on OR sequence analysis[22,23] and numerous site-directed mutagenesis data,[23,24] which are located in the upper portion of TM3 and TM5−TM7. Here, we highlight 4 residues in TM2 near a conserved allosteric site (centered at $D^{2.50}$). The allosteric site in nonolfactory class A GPCRs (typically composed of $D^{2.50}$, $N^{3.35}$, and $S^{3.39}$) is known to bind the Na^+ ion, which modulates the receptors' activation and affinity/response to ligands (reviewed in ref 29). Most ORs contain a second acidic residue ($E^{3.39}$) at this site, which might also accommodate divalent cations.[29] While copper ions play important roles in the recognition of sulfur odorants,[30,31] it remains unclear whether this conserved site in the ORs is involved. The machine learning model established here outperformed existing models using full sequences.[17,19] The pocket residues are essential for understanding how chemically similar odorants are differentiated by the OR family with such high specificity/selectivity.

So far, research focusing on specific OR−ligand recognition has mostly employed molecular modeling (e.g., homology modeling, docking, and molecular simulations) verified by site-directed mutagenesis and functional assays of individual ORs, such as the studies reviewed in ref 32, as well as the more recent work on hOR1A1 for R-/S-carvone enantiomers,[33] hOR5AN1 and mOR215-1 for musk odorants,[34] zebrafish ORs for bile acids/salts,[35] and a virtual screening for new mOR-EG ligands.[36] This approach provides valuable insights into OR−ligand recognition and will continue to generate data for new ORs and ligands. Since it relies on experimental data to generate predictive molecular models, this approach is not suitable for large-scale OR−ligand pairing. The molecular modeling process can be automated to enable large-scale studies;[37] however, the performance has yet to be tested. Ligand QSAR/SAR models using machine learning have also been adopted to predict new OR ligands.[38,39] This approach allows a rapid virtual screening of large compound databases and is widely used in drug design and drug toxicity prediction.[40] It is limited to the target receptor and the chemical scaffolds of the known ligands. However, the application on ORs will gradually enrich ligand data and reduce the bottleneck of our PCM model.

The machine learning PCM approach established here is readily applicable to the entire mammalian OR family. It will significantly accelerate OR−ligand mapping and OR deorphanization. It is an open loop process where newly identified OR−odorant pairs can be added to continuously improve the model. Because we optimized the model to maximize the hit rate (to reduce the cost of *in vitro* assays), this consequently gave way to false negatives (Figure S4C). Therefore, repeating the prediction−test loop is necessary to rescue the false negatives by injecting new training data. Note that the lack of response of many orphan ORs might be due to impaired functions in heterologous cells, e.g., lack of cell surface expression.[41] For instance, ∼30% of the mORs responding to acetophenone *in vivo* did not show significant responses in heterologous cells.[18] Such cases may be present in the nonresponsive ORs in the *in vitro* test set, the proportion of which is difficult to estimate.

This approach is mostly applicable to large protein families like GPCRs or promiscuous proteins, such as functionally related enzymes,[34] odorant/pheromone-binding proteins in insects,[35] intrinsically disordered protein regions,[36] as well as GPCR-G protein binding partners.[37] The approach focuses on the sequence of the binding region, which overcomes the difficulties in obtaining high-resolution structures or full sequence alignments. It may find applications in, for example, predicting off-target activities in drug design, targeting insect pheromone receptors for pest control, or studies of protein−protein interactions and protein evolution. It requires sequence alignment and a number of known ligands as input data. The selection of relevant residues is important, which enables knowledge-based human intervention to reduce the dimensionality and enhance machine learning on scarce data. Combining *in vitro* functional assays, site-directed mutagenesis, knowledge of GPCR structures and sequences, as well as molecular modeling, we could generate heuristics to decipher how nature has encoded the specific functions of ORs into their varied sequences.

The model is currently limited to the transmembrane domain where the sequence alignment has been established. The loop regions may be addressed for OR subfamilies for which good sequence alignments can be obtained. The discovery of residue subsets associated with given functions could indicate evolutionary hotspots and compensate for existing tools such as phylogenetic analysis based on full sequences.

MATERIALS AND METHODS

Chemicals and OR Constructs. Odorants were purchased from Sigma-Aldrich. They were dissolved in DMSO to make stock solutions at 1 mM and then freshly diluted in optimal MEM (ThermoFisher) to prepare the odorant stimuli. The OR constructs were kindly provided by Dr. Hanyi Zhuang (Shanghai Jiaotong University, China). Site-directed mutants were constructed using the Quikchange site-directed mutagenesis kit (Agilent Technologies). The sequences of all plasmid constructs were verified by both forward and reverse sequencing (Sangon Biotech, Shanghai, China). The list of primers used in this study are listed in Table S9.

Cell Culture and Transfection. We used Hana3A cells, a HEK293T-derived cell line that stably expresses receptor-transporting proteins (RTP1L and RTP2), receptor expression-enhancing protein 1 (REEP1), and olfactory G protein ($G\alpha_{olf}$).[42] The cells were grown in MEM (Corning) supplemented with 10% (v/v) fetal bovine serum (FBS; ThermoFisher) and 100 μg/mL penicillin−streptomycin (ThermoFisher), 1.25 μg/mL amphotericin (Sigma-Aldrich), and 1 μg/mL puromycin (Sigma-Aldrich).

All constructs were transfected into the cells using Lipofectamine 2000 (ThermoFisher). Before the transfection, the cells were plated on 96-well plates (NEST) and incubated overnight in MEM with 10% FBS at 37 °C and 5% CO_2. For each 96-well plate, 2.4 μg of pRL-SV40, 2.4 μg of CRE-Luc, 2.4 μg of mouse RTP1S, and 12 μg of receptor plasmid DNA were transfected. The cells were subjected to a luciferase assay 24 h after transfection.

Luciferase Assay. The luciferase assay was performed with the Dual-Glo luciferase assay kit (Promega) following the protocol in ref 42. OR activation triggers the $G\alpha_{olf}$-driven AC-cAMP-PKA signaling cascade and phosphorylates CREB. Activated CREB induces luciferase gene expression, which can be quantified luminometrically [measured here with a bioluminescence plate reader (MD SPECTRAMAX L)]. Cells were cotransfected with firefly and *Renilla* luciferases where firefly luciferase served as the cAMP reporter. *Renilla* luciferase is driven by a constitutively active simian virus 40 (SV40) promoter (pRL-SV40; Promega), which served as a control for cell viability and transfection efficiency. The ratio between firefly luciferase versus *Renilla* luciferase was measured. Normalized OR activity was calculated as $(L_N - L_{min})/(L_{max} - L_{min})$, where L_N is the luminescence in response to the odorant, and L_{min} and L_{max} are the minimum and maximum luminescence values on a plate, respectively. The assay was carried out as follows: 24 h after transfection, the medium was replaced with 100 μL of odorant solution (at different doses) diluted in optimal MEM (ThermoFisher), and cells were further incubated for 4 h at 37 °C and 5% CO_2. After incubation in lysis buffer for 15 min, 20 μL of Dual-Glo luciferase reagent was added to each well of a 96-well plate, and firefly luciferase luminescence was measured. Next, 20 μL of Stop-Glo luciferase reagent was added to each well, and *Renilla* luciferase luminescence was measured. The data analysis followed the published procedure in ref 42. Three-parameter dose−response curves were fitted with GraphPad Prism 8.

Molecular Modeling. Homology models of mOR256-3, mOR256-8, and mOR256-31 were built using the approach in our previous work.[24,27] Four X-ray crystal structures of class A GPCRs were used as templates, rhodopsin (1U19), CXCR4 (3ODU), A2aR (2YDV), and CXCR1 (2LNL), to build 100 models with Modeler v9.15.[43] For docking, we chose the model with the lowest DOPE score. Autodock Vina[44] and the Haddock 2.2 Web server[45] were used to identify a common top-ranked binding pose for each odorant. Residues in the putative ligand-binding pocket were set flexible during docking. Enhanced-sampling all-atom molecular dynamics simulations were performed in a bilayer of an explicit POPC membrane (see the Methods section in the SI for details). A cluster analysis of the ligand-binding pose was carried out on the simulation trajectories using the Gromacs Cluster tool. The middle structure of the most populated cluster was selected as the final binding pose.

Proteochemometric Machine Learning Model. We assembled the response data of 720 ORs and 244 odorants from the literature to construct the training set (Data File S1). Ambiguous data records (i.e., OR responses without clear dose-dependent data) were discarded. The full training set contained 1293 responsive OR−odorant pairs (composed of 392 ORs and 244 odorants) and 14 459 OR−odorant pairs that have been reported to be nonresponsive *in vitro* (composed of 550 ORs and 127 odorants, including 318 orphan ORs). Each OR−odorant pair was represented by a vector composed of physicochemical descriptors (features) of the OR sequence and the odorant (see the Methods section in the SI for details). The OR−odorant pairs in the training set were labeled "positive" or "negative" according to the response data for supervised machine learning. The test set was constructed in the same manner without labels. The test set contained 360 ORs (including 346 orphan ORs) available in our laboratory, paired with the 7 odorants tested in this study. RF and SVM classification models were built with the Caret package in R.[46] RF performed better than SVM and was chosen for the final model. The R code generated during this study is available as a Jupyter notebook, along with the input and output data, at https://github.com/chemosim-lab/OlfactoryReceptors under the GNU General Public License v3.0. The Jupyter notebook illustrates step-by-step the model building, training, and the *in vitro* assessment. The process is illustrated in Figure S2A. More details can be found in the Methods section in the SI.

Safety Statement. No unexpected or unusually high safety hazards were encountered.

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acscentsci.1c01495.

Supplementary Methods section and additional data and figures including the dose-dependent responses of mOR256-31 mutants, workflow and conservation of *poc60* residues, *in vitro* test sets, and an analysis of the RF classifiers' predictivity on OR response to odorants (PDF)

Data File S1 (XLSX)

Data File S2 (XLSX)

AUTHOR INFORMATION

Corresponding Authors

Xiaojing Cong − *Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice 06108, France;* Present Address: X.C.: Institut de Génomique Fonctionnelle, University of Montpellier, CNRS, INSERM, 34094 Montpellier, France; orcid.org/0000-0002-5051-2392; Email: xiaojing.cong@igf.cnrs.fr

Yiqun Yu − *Ear, Nose & Throat Institute, Department of Otolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Clinical and Research Center for Olfactory Disorders, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai*

200031, People's Republic of China; Email: yu_yiqun@fudan.edu.cn

Jérôme Golebiowski − *Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice 06108, France; Department of Brain and Cognitive Sciences, Daegu Gyeongbuk Institute of Science and Technology, Daegu 711-873, South Korea*; orcid.org/0000-0002-3675-1952; Email: jerome.golebiowski@gmail.com

Authors

Wenwen Ren − *Institutes of Biomedical Sciences, Fudan University, Shanghai 200031, People's Republic of China*

Jody Pacalon − *Université Côte d'Azur, CNRS, Institut de Chimie de Nice UMR7272, Nice 06108, France*

Rui Xu − *School of Life Sciences, Shanghai University, Shanghai 200444, People's Republic of China*

Lun Xu − *Ear, Nose & Throat Institute, Department of Otolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China*

Xuewen Li − *School of Life Sciences, Shanghai University, Shanghai 200444, People's Republic of China*

Claire A. de March − *Department of Molecular Genetics and Microbiology, and Department of Neurobiology, and Duke Institute for Brain Sciences, Duke University Medical Center, Durham, North Carolina 27710, United States*

Hiroaki Matsunami − *Department of Molecular Genetics and Microbiology, and Department of Neurobiology, and Duke Institute for Brain Sciences, Duke University Medical Center, Durham, North Carolina 27710, United States*

Hongmeng Yu − *Ear, Nose & Throat Institute, Department of Otolaryngology, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Clinical and Research Center for Olfactory Disorders, Eye, Ear, Nose & Throat Hospital, Fudan University, Shanghai 200031, People's Republic of China; Research Units of New Technologies of Endoscopic Surgery in Skull Base Tumor, Chinese Academy of Medical Sciences, Beijing 100730, People's Republic of China*

Complete contact information is available at:
https://pubs.acs.org/10.1021/acscentsci.1c01495

Author Contributions

◆X.C. and W.R. contributed equally. X.C., H.M., Y.Y., and J.G. designed the research. X.C. and J.P. collected and analyzed literature data. X.C., J.P., and J.G. performed and analyzed *in silico* experiments. W.R., R.X., L.X., X.L., C.A.d.M., and Y.Y. performed *in vitro* experiments. W.R., R.X., L.X., H.Y., C.A.d.M., H.M., and Y.Y. analyzed *in vitro* data. X.C., Y.Y., and J.G. wrote the paper. All authors have given approval to the final version of the manuscript.

Notes

The authors declare the following competing financial interest(s): H.M. has received royalties from ChemCom, research grants from Givaudan, and consultant fees from Kao.

■ ACKNOWLEDGMENTS

This work received funding from the French National Research Agency (ANR, grant NEUROLF to J.G.) as part of the ANR-NSF-NIH Collaborative Research in Computational Neuroscience; the French government, through the UCA-Jedi project managed by the ANR (ANR-15-IDEX-01 to X.C. and J.G.) and, in particular, by the interdisciplinary Institute for Modeling in Neuroscience and Cognition (NeuroMod) of the Université Côte d'Azur; the German Research Foundation (DFG, grant CO 1715/1-1 to X.C.); the Roudnitska Foundation (France, a PhD fellowship to J.P.); GIRACT (Switzerland, a research fellowship to J.P.); the National Institutes of Health (NIH, grants DC016224 to H.M. and K99/R00DC018333 to C.D.M.); the National Science Foundation (NSF, grant 1515801 to H.M.); the National Natural Science Foundation of China (Grants 32070996 and 31771155 to Y.Y., 31900714 to W.R.); Science and Technology Commission of Shanghai Municipality (grant 21140900600 to Y.Y.); Shanghai Municipal Human Resources and Social Security Bureau (Shanghai Talent Development Fund 2018112 to Y.Y.); Eye, Ear, Nose and Throat Hospital, Fudan University (Excellent Doctors-Excellent Clinical Researchers Program SYB202002 to Y.Y.); and the New Technologies of Endoscopic Surgery in Skull Base Tumor: CAMS Innovation Fund for Medical Sciences (CIFMS; 2019-I2M-5-003 to H.Y.). This work was granted access to the HPC resources of CINES under the allocation 2018−2019 A0040710477 made by GENCI. We also thank Dr. Sébastien Fiorucci, Dr. Jérémie Topin, Dr. Cédric Bouysset, Priyanka Meesa, and Shiyi Jiang for critically reading and editing the manuscript.

■ REFERENCES

(1) Santos, R.; Ursu, O.; Gaulton, A.; Bento, A. P.; Donadi, R. S.; Bologa, C. G.; Karlsson, A.; Al-Lazikani, B.; Hersey, A.; Oprea, T. I.; et al. A comprehensive map of molecular drug targets. *Nat. Rev. Drug Discovery* **2017**, *16* (1), 19−34.

(2) Fredriksson, R.; Lagerstrom, M. C.; Lundin, L. G.; Schioth, H. B. The G-protein-coupled receptors in the human genome form five main families. Phylogenetic analysis, paralogon groups, and fingerprints. *Mol. Pharmacol.* **2003**, *63* (6), 1256−72.

(3) Bushdid, C.; Magnasco, M. O.; Vosshall, L. B.; Keller, A. Humans can discriminate more than 1 trillion olfactory stimuli. *Science* **2014**, *343* (6177), 1370−2.

(4) Cong, X.; Topin, J.; Golebiowski, J. Class A GPCRs: Structure, Function, Modeling and Structure-based Ligand Design. *Curr. Pharm. Des* **2017**, *23* (29), 4390−4409.

(5) Venkatakrishnan, A. J.; Deupi, X.; Lebon, G.; Tate, C. G.; Schertler, G. F.; Babu, M. M. Molecular signatures of G-protein-coupled receptors. *Nature* **2013**, *494* (7436), 185−94.

(6) Malnic, B.; Hirono, J.; Sato, T.; Buck, L. P. Combinatorial receptor codes for odors. *Cell* **1999**, *96*, 713−723.

(7) Tcatchoff, L.; Nespoulous, C.; Pernollet, J. C.; Briand, L. A single lysyl residue defines the binding specificity of a human odorant-binding protein for aldehydes. *FEBS Lett.* **2006**, *580* (8), 2102−8.

(8) Briand, L.; Eloit, C.; Nespoulous, C.; Bezirard, V.; Huet, J. C.; Henry, C.; Blon, F.; Trotier, D.; Pernollet, J. C. Evidence of an odorant-binding protein in the human olfactory mucus: location, structural characterization, and odorant-binding properties. *Biochemistry* **2002**, *41* (23), 7241−52.

(9) Ferrer, I.; Garcia-Esparcia, P.; Carmona, M.; Carro, E.; Aronica, E.; Kovacs, G. G.; Grison, A.; Gustincich, S. Olfactory receptors in non-chemosensory organs: the nervous system in health and disease. *Front Aging Neurosci* **2016**, *8*, 163.

(10) Kang, N.; Koo, J. Olfactory receptors in non-chemosensory tissues. *BMB reports* **2012**, *45* (11), 612−22.

(11) Kang, N.; Kim, H.; Jae, Y.; Lee, N.; Ku, C. R.; Margolis, F.; Lee, E. J.; Bahk, Y. Y.; Kim, M. S.; Koo, J. Olfactory marker protein expression is an indicator of olfactory receptor-associated events in non-olfactory tissues. *PloS one* **2015**, *10* (1), No. e0116097.

(12) Lee, S. J.; Depoortere, I.; Hatt, H. Therapeutic potential of ectopic olfactory and taste receptors. *Nat. Rev. Drug Discov* **2019**, *18* (2), 116−138.

(13) Lapinsh, M.; Prusis, P.; Gutcaits, A.; Lundstedt, T.; Wikberg, J. E. Development of proteo-chemometrics: a novel technology for the analysis of drug-receptor interactions. *Biochim. Biophys. Acta* **2001**, *1525* (1−2), 180−90.

(14) Rifaioglu, A. S.; Atas, H.; Martin, M. J.; Cetin-Atalay, R.; Atalay, V.; Dogan, T. Recent applications of deep learning and machine intelligence on in silico drug discovery: methods, tools and databases. *Brief Bioinform* **2019**, *20* (5), 1878−1912.

(15) Mousavian, Z.; Masoudi-Nejad, A. Drug-target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin Drug Metab Toxicol* **2014**, *10* (9), 1273−87.

(16) Ezzat, A.; Wu, M.; Li, X. L.; Kwoh, C. K. Computational prediction of drug-target interactions using chemogenomic approaches: an empirical survey. *Brief Bioinform* **2019**, *20* (4), 1337−1357.

(17) Liu, X.; Su, X.; Wang, F.; Huang, Z.; Wang, Q.; Li, Z.; Zhang, R.; Wu, L.; Pan, Y.; Chen, Y.; et al. ODORactor: a web server for deciphering olfactory coding. *Bioinformatics* **2011**, *27* (16), 2302−3.

(18) Jiang, Y.; Gong, N. N.; Hu, X. S.; Ni, M. J.; Pasi, R.; Matsunami, H. Molecular profiling of activated olfactory neurons identifies odorant receptors for odors in vivo. *Nat. Neurosci* **2015**, *18* (10), 1446−54.

(19) Saito, H.; Chi, Q.; Zhuang, H.; Matsunami, H.; Mainland, J. D. Odor coding by a Mammalian receptor repertoire. *Sci. Signal* **2009**, *2* (60), No. ra9.

(20) Chepurwar, S.; Gupta, A.; Haddad, R.; Gupta, N. Sequence-Based Prediction of Olfactory Receptor Responses. *Chem. Senses* **2019**, *44* (9), 693−703.

(21) Chen, K. M.; Keri, D.; Barth, P. Computational design of G Protein-Coupled Receptor allosteric signal transductions. *Nat. Chem. Biol.* **2020**, *16* (1), 77−86.

(22) Man, O.; Gilad, Y.; Lancet, D. Prediction of the odorant binding site of olfactory receptor proteins by human-mouse comparisons. *Protein Sci.* **2004**, *13* (1), 240−54.

(23) de March, C. A.; Kim, S. K.; Antonczak, S.; Goddard, W. A., 3rd; Golebiowski, J. G protein-coupled odorant receptors: From sequence to structure. *Protein Sci.* **2015**, *24* (9), 1543−8.

(24) de March, C. A.; Topin, J.; Bruguera, E.; Novikov, G.; Ikegami, K.; Matsunami, H.; Golebiowski, J. Odorant Receptor 7D4 Activation Dynamics. *Angew. Chem.* **2018**, *57* (17), 4554−4558.

(25) Yu, Y.; de March, C. A.; Ni, M. J.; Adipietro, K. A.; Golebiowski, J.; Matsunami, H.; Ma, M. Responsiveness of G protein-coupled odorant receptors is partially attributed to the activation mechanism. *Proc. Natl. Acad. Sci. U. S. A.* **2015**, *112* (48), 14966−71.

(26) Nara, K.; Saraiva, L. R.; Ye, X.; Buck, L. B. A large-scale analysis of odor coding in the olfactory epithelium. *J. Neurosci.* **2011**, *31* (25), 9179−91.

(27) de March, C. A.; Yu, Y.; Ni, M. J.; Adipietro, K. A.; Matsunami, H.; Ma, M.; Golebiowski, J. Conserved Residues Control Activation of Mammalian G Protein-Coupled Odorant Receptors. *J. Am. Chem. Soc.* **2015**, *137* (26), 8611−8616.

(28) Matthews, B. W. Comparison of the predicted and observed secondary structure of T4 phage lysozyme. *Biochim Biophys Acta - Protein Struc* **1975**, *405* (2), 442−451.

(29) Katritch, V.; Fenalti, G.; Abola, E. E.; Roth, B. L.; Cherezov, V.; Stevens, R. C. Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* **2014**, *39* (5), 233−44.

(30) Duan, X.; Block, E.; Li, Z.; Connelly, T.; Zhang, J.; Huang, Z.; Su, X.; Pan, Y.; Wu, L.; Chi, Q.; et al. Crucial role of copper in detection of metal-coordinating odorants. *Proc. Natl. Acad. Sci. U. S. A.* **2012**, *109* (9), 3492−7.

(31) Wang, J.; Luthey-Schulten, Z. A.; Suslick, K. S. Is the olfactory receptor a metalloprotein? *Proc. Natl. Acad. Sci. U. S. A.* **2003**, *100* (6), 3035−9.

(32) Launay, G.; Sanz, G.; Pajot-Augy, E.; Gibrat, J. F. Modeling of mammalian olfactory receptors and docking of odorants. *Biophys Rev.* **2012**, *4* (3), 255−269.

(33) Geithe, C.; Protze, J.; Kreuchwig, F.; Krause, G.; Krautwurst, D. Structural determinants of a conserved enantiomer-selective carvone binding pocket in the human odorant receptor OR1A1. *Cell. Mol. Life Sci.* **2017**, *74* (22), 4209−4229.

(34) Sato-Akuhara, N.; Horio, N.; Kato-Namba, A.; Yoshikawa, K.; Niimura, Y.; Ihara, S.; Shirasu, M.; Touhara, K. Ligand Specificity and Evolution of Mammalian Musk Odor Receptors: Effect of Single Receptor Deletion on Odor Detection. *J. Neurosci.* **2016**, *36* (16), 4482−91.

(35) Cong, X.; Zheng, Q.; Ren, W.; Cheron, J. B.; Fiorucci, S.; Wen, T.; Zhang, C.; Yu, H.; Golebiowski, J.; Yu, Y. Zebrafish olfactory receptors ORAs differentially detect bile acids and bile salts. *J. Biol. Chem.* **2019**, *294* (17), 6762−6771.

(36) Yuan, S.; Dahoun, T.; Brugarolas, M.; Pick, H.; Filipek, S.; Vogel, H. Computational modeling of the olfactory receptor Olfr73 suggests a molecular basis for low potency of olfactory receptor-activating compounds. *Commun. Biol.* **2019**, *2*, 141.

(37) Launay, G.; Teletchea, S.; Wade, F.; Pajot-Augy, E.; Gibrat, J. F.; Sanz, G. Automatic modeling of mammalian olfactory receptors and docking of odorants. *Protein Eng. Des Sel* **2012**, *25* (8), 377−86.

(38) Caballero-Vidal, G.; Bouysset, C.; Grunig, H.; Fiorucci, S.; Montagne, N.; Golebiowski, J.; Jacquin-Joly, E. Machine learning decodes chemical features to identify novel agonists of a moth odorant receptor. *Sci. Rep* **2020**, *10* (1), 1655.

(39) Bushdid, C.; de March, C. A.; Fiorucci, S.; Matsunami, H.; Golebiowski, J. Agonists of G-Protein-Coupled Odorant Receptors Are Predicted from Chemical Features. *J. Phys. Chem. Lett.* **2018**, *9* (9), 2235−2240.

(40) Muratov, E. N.; Bajorath, J.; Sheridan, R. P.; Tetko, I. V.; Filimonov, D.; Poroikov, V.; Oprea, T. I.; Baskin, II; Varnek, A.; Roitberg, A.; et al. QSAR without borders. *Chem. Soc. Rev.* **2020**, *49* (11), 3525−3564.

(41) Ikegami, K.; de March, C. A.; Nagai, M. H.; Ghosh, S.; Do, M.; Sharma, R.; Bruguera, E. S.; Lu, Y. E.; Fukutani, Y.; Vaidehi, N.; et al. Structural instability and divergence from conserved residues underlie intracellular retention of mammalian odorant receptors. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (6), 2957−2967.

(42) Zhuang, H.; Matsunami, H. Evaluating cell-surface expression and measuring activation of mammalian odorant receptors in heterologous cells. *Nat. Protoc* **2008**, *3* (9), 1402−13.

(43) Eswar, N.; Webb, B.; Marti-Renom, M. A.; Madhusudhan, M. S.; Eramian, D.; Shen, M. Y.; Pieper, U.; Sali, A. Comparative protein structure modeling using Modeller. *Curr. Protoc. Bioinform.* **2006**, *15* (1), 5.6.1.

(44) Trott, O.; Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **2009**, *31* (2), 455−461.

(45) van Zundert, G. C. P.; Rodrigues, J.; Trellet, M.; Schmitz, C.; Kastritis, P. L.; Karaca, E.; Melquiond, A. S. J.; van Dijk, M.; de Vries, S. J.; Bonvin, A. The HADDOCK2.2 Web Server: User-Friendly Integrative Modeling of Biomolecular Complexes. *J. Mol. Biol.* **2016**, *428* (4), 720−725.

(46) Kuhn, M. Building Predictive Models in R Using the caret Package. *J. Stat Softw* **2008**, *28* (5), 26.